DOCUMENT RESUME

ED 036 884 EA 002 721

AUTHOR Welty, Gordon
TITLE Evaluation of Public School Programs.
INSTITUTION Pittsburgh Public Schools, Pa. Office of Research.
PUB DATE 9 Jul 69
NOTE 14p.; Paper presented at L.R.D.C. (November 18, 1968)

EDRS PRICE EDRS Price MF-$0.25 HC-$0.80
DESCRIPTORS Criteria, *Evaluation Techniques, Feedback, Input Output Analysis, *Program Design, *Program Evaluation, *Public Schools, Standards, Student Characteristics, Theories

ABSTRACT

This article outlines the procedures followed in program evaluation in Pittsburgh public schools. A program design is obtained by asking the field staff a series of specific questions. As the staff interact, problems about the program are resolved. The consensus achieved is the basis for standardization of activities in the field. The next step in evaluation is a panel meeting that brings expert criticism to bear on the program design. When the implicit theory is criticized and the structure is compatible with the design criteria, the information is given back to the program manager. The second stage involves compatibility testing to pinpoint conflicts and congruence testing of actual activity and the program design. (MF)

ED036884

EVALUATION OF PUBLIC SCHOOL PROGRAMS

Gordon Welty
Chatham College

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
OFFICE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION POSITION OR POLICY.

Presented at L. R. D. C.
18 Nov. 1968

EA 002 721

SUBJECT: Evaluation of Public School Programs

AUTHOR: Gordon Welty

DATE SUBMITTED: July 9, 1969

OFFICE OF RESEARCH

PITTSBURGH PUBLIC SCHOOLS

Malcolm M. Provus, Director

I want to emphasize that the activities which we will discuss here, although presented serially, in fact occur concurrently. I also want to emphasize that the object of evaluation, as we see it, is to compare performance with a standard, and provide feedback on discrepancies. This feedback permits decision-makers to change either behavior or the standard, and thus equalize the two.

As Mrs. McBroom pointed out, the first thing needed in an evaluation is a program design or blueprint. This design tells us what it is we're evaluating, what we can expect to find out in the field. First, I will talk about gathering information for this blueprint.

The traditional way of determining what the program is, was to get a copy of the funding proposal and say "oh, this is the design, let's go see if this is happening." Anyone who has done educational evaluations knows this is not a good procedure. The funding proposal has little to do with what's going on in the field. A funding proposal is designed to get money, not provide a blueprint for a program. So we must look elsewhere.

In Pittsburgh we've chosen the people who are actually doing the field work, the teachers, librarians, and so forth as the source of the program design. In essence we ask them "What are you trying to do here?" and when they tell us, we write it down. This is the program design.

For instance, in Pittsburgh we have been evaluating an instrumental music program. The type of questions we have asked program staff included:

1. What characteristics does the student have on entering the program that you wish to change before he leaves the program? How do you want them to change?

For example, what kinds of musical ability do you want to develop? What attitudes do you want the child to have toward himself, toward music, toward school? What changes, if any, in his personal characteristics do you want to foster?

2. What characteristics must the student have in order to enter the program? Are there specific musical abilities he must have? Specific personal characteristics? Do his academic grades have to be kept at a certain level?

3. What new skills does the teacher develop as a result of the program? What old skills does he improve?

4. What specific materials are required by the program? By this we mean such things as method books, music stands, etc. down to extra E strings and repair request forms. Who chooses the materials used?

5. What facilities are necessary to the program? For example, is there a minimum size for the room in which lessons are taught? Minimum acoustical properties?

The first question refers to student variables. The third question refers to staff variables. The second, fourth and fifth questions refer to preconditions for program operation.*

A couple of years ago, Esther Kresh evaluated the Team Teaching Program in Pittsburgh and reported there were 131 different programs.** There was a different program for each team! This is what we want to avoid. We need <u>one</u> program design, and we can use this as a means of achieving unanimity in the field. If the 131 different teaching teams had been brought together, it would have been obvious to all of them that they had no real program. If there is more than one blueprint, there will not be a building. This unique blueprint we seek is the program design.

So what we've been doing in Pittsburgh has been to assemble the various teachers and managers in one spot. We've tried heterogeneous

---

*I would like to thank Laurie Dancy who is evaluating the Instrumental Music Program for this information.

**Esther Kresh and Russell Scott. "Team Teaching Program--1967 Report" (Pittsburgh Public Schools: Pittsburgh, n.d.).

and homogeneous grouping. We've assembled the whole group and samples of the group. Then we ask them a series of very specific questions. We might ask, in a remedial reading program, "Do you wish to change reading achievement or do you wish to make a diagnosis of reading difficulties." The answers to these questions make up the program design. The first time around, the design is obviously going to be vague and ambiguous. The teachers, project managers, and other staff employ the usual educational cliches. But by the second or third time around, more precision is attained.

We see that we can assemble information for the preparation of a program design by questioning program staff. A second function of the meeting, independent of information gathering that I've just discussed, is the function of consensus building.

We want one and only one program in the field to correspond to our one and only one program design. Thus the teachers and project managers and other staff must function as a collective. All of the teachers and managers must internalize the program. They must internalize the concepts as defined in the meetings. You may wonder what we mean by consensus and internalizing of the program.

For our purposes, consensus is the establishment of a working agreement, and is defined as a minimization of variance of rank-ordered objectives. If we have three items, representing three objectives, with the items listed as rows, and possible ranks as columns, we have a square matrix with the rater response entered in the cells as x's and o's for two hypothetical raters,

RANK

		1	2	3
	1	x		
Item	2		x	
	3			x

Here variance in ranking is minimal. This is consensus. A working agreement has been achieved.

RANK

		1	2	3
	1			x
Item	2	x		
	3		x	

Here too we have consensus.

RANK

		1	2	3
	1	x		o
Item	2	o	x	
	3		o	x

Here we have dissensus. These arrays have proved convenient ways of discovering values to which staff subscribe. Of course, these are all behavioral definitions: in this context "consensus" does not mean a transcendental entity.

As the staff interact together at the meetings, problems and conflicts surface and can be worked out. As these problems about the program are resolved, consensus is achieved. Then the field operations will be able to come into accord with the one and only one program design.

So we have examined two functions of these meetings: to provide information and to generate consensus. The information generated is

written up into the program design, and the consensus which is achieved is the basis for standardization of activities in the field. These functions are concurrently filled. The first could be conceived as a cognitive activity, the second as an affective activity, of evaluation.

Now we have the program design, crude as it presently is, in hand. The question immediately arises: Is this blueprint a good blueprint? As we have pointed out, evaluation is comparison, with negative or discrepancy reports facilitating program improvement. Thus we want to improve the program design. The answer to the question "Is this a good blueprint?" is sought at what we call panel meetings. A panel is a mechanism for bringing expert criticism to bear on the program design.

The program design may be theoretically sound and structurally unsound, or vice-versa. For instance, a theoretically sound remedial program might employ a certain learning program such as the Sullivan materials. It may be structurally unsound because the dimension of staff qualifications is passed over by the staff as unimportant.*

Conversely, the program may have a (woefully) deficient theory, as may well be the case with a Team Teaching Program, yet if all the major dimensions of the program are specified, then we would say the design of the program was structurally, but not theoretically, sound.

So we want to examine two aspects of the program design, the implicit theory and the structure. To examine the theory, we bring in

---

*On the dimensional approach, cf. A. . Melton's article "Learning" in the Encyclopedia of Educational Research, ed. W.S. Monroe (N.Y.: Macmillan, 1941), pp. 667-686.

a specialist in the substantive area of the program. This is an expert who will examine the design and say, for instance, "you haven't allotted enough time for this rote learning activity. You need at least 10 minutes a day practice for mastery." These problems are recorded as problems of the design. Now I'll discuss the nature of design criteria and the structure of the program blueprint.

Unlike the method for examining the theory, where we use an expert to analyse the design structurally, we must compare it with a set of generalized design criteria.* We can conceive a program as consisting of inputs which go through some process and give us outputs. First we will consider inputs. To characterize inputs we have three things, <u>variables</u> which might consist of student performance measures, staff measures, indeed anything which is to vary as a result of the program. We also have <u>preconditions</u> which further describe students, staff, and other necessities or overhead items. These do not vary through the program. Thus the difference between precondition and variable is that the variable can be changed by the program, the precondition cannot. For instance, a measure of reading achievement could be a variable; a measure of I.Q. would be a precondition.

The third category we see under inputs is <u>criteria</u>. The criteria specify ranges or values of our preconditions and variables. Specifically, the criteria on Student measures and Student conditions represent the selection criteria of a program.

For instance, a remedial reading program might specify that the

---

*See "Design Criteria" following.

students have an I.Q. above 85, so they can benefit from the remediation. This means that as a precondition you will have intelligence as measured by some standardized I.Q. test. Here the student characteristic would be bounded below by the criterion that I.Q. must be above 85.

A program might specify that the student be in the third grade to participate. Here the preconditon is grade in schcol, and the criterion specifies third grade. If the student is in the second or in the fourth grade he's not supposed to be in the program.

The variable, alternatively, could be reading achievement. Here you could say that to be in a remedial reading program, performance must be at least one year below grade level. Reading achievement is the variable; more than one grade level deficient, the criterion on the variable. Staff measures can be change variables in the case where a training program exists within a larger program. Moving along the continuum we next come to process.

Under process we again have variables.* These would include student activities and could state that the student reads the Sullivan materials. Staff activities in a remedial reading program would state that the teacher's function is to provide positive reinforcement for students who are reading the Sullivan materials.

Now we turn to criteria. For instance, on student activities it might be specified that each student is to spend 80 percent of his time reading Sullivan materials. The teacher is to spend 90 percent of her time positively reinforcing the child who is using the Sullivan materials. There must be sufficient conditions for transforming the input variables from their initial value into the terminal or exit

---

*On "process" cf. Melton, op.cit., p. 667.

value of the output variables. So we have finally come to outputs.

With output variables, we have the same things as we had under input variables. In a remedial reading program we would have reading achievement as a variable, and preconditions would remain the same. In the case of outputs, the criteria specify the goals of the program in terms of the variable. For instance, a goal could be specified by the criterion that reading achievement be at grade level.

It is of course possible that reading achievement is not brought to grade level. Student's reading may stay at the same one year deficient level the whole way through the program. At the end of the year he violates the precondition of being in the third grade, and he's eliminated from the program. Obviously, success has not been achieved in this case.

Let us summarize our discussion of design criteria by looking at the kinds of problems we can uncover by systematic comparison of blueprint with the generalized design criteria. In terms of the design criteria we look at the program and ascertain that a preprimary program requires teachers' aides. First we ask "Is this a programmed part of the project or is this an ad hoc part of the project?" If it's a programmed part, staff qualifications must tell you what it means to be a teacher's aide. Under process variables, you must be able to find out what are the activities of a teacher's aide.

If the evaluator doesn't find these items, if he finds, as is usually the case, that a teacher's aide is provided for in the project, but it doesn't say anywhere what the teacher's aide is supposed to do, who she is, what her qualifications are, and so forth, then the evaluator knows that there is a deficiency in the program with re-

DESIGN CRITERIA

Inputs	Process	Outputs
I. Variables--the things the program is attempting to change A. Student Variables B. Staff Variables C. Other Variables	Variables--those activities which change inputs into desired outputs A. Student Activities B. Staff Activities 1. Functions and Duties 2. Communication a. Intra-staff b. With Others	Variables--the changes that have come about A. Student Variables B. Staff Variables C. Other Variables
II. Preconditions--the things that are prerequisite to program operation yet remain constant throughout the program Student Conditions Staff Qualifications Administrative Support Media Facilities Time		Preconditions--same throughout the program
III. Criteria must be specified for each input variable and precondition above. The criteria specified for student variables and preconditions constitute the selection criteria of the program.	Criteria must be specified for each of the process variables.	Criteria are specified on the variables to define the goals of the program. The participant is released from the program if he achieves the goal of the program or if he violates a precondition.

gard to the definition of the teacher's aide. And he must point this out. In the absence of such information, it is not possible to know where the program has not been implemented, and thus not possible to use product data findings for program change and improvement.

So we see that design criteria enable us to explicate the structure of the program design and to facilitate valid measurement of process and product. This approach is similar to the functional analysis of program planning and budgeting, where each function is broken down into other smaller units, always under the criterion of sufficiency to realize the larger functions. As you can see, with regard to staff activity under process, if the staff activity specifies that teachers positively reinforce students reading, we could take just this function to the level of a whole program. This is exactly what is done when one evaluates inservice training.

So we can keep pulling subprograms from the process area of the design, make these into complete programs, and break them down in turn into further subprograms.

Thus we see we can look at the implicit theory or the structure of the program designs we have. In both cases we are criticizing the blueprint, in the one case by a comparison of the implicit theory with the expert's knowledge of the substance area. This is to provide for theoretical meaningfulness. In the other case, we compare the blueprint with a set of generalized design criteria. Indeed, this latter comparison is to guarantee dimensional homogeneity, without which the design becomes methodologically quite literal nonsense. Both of these functions enable us to rectify the program design we have. When this is done, when the information is gathered and consensus about the design is generated, when the implicit theory is criticized and the struc-

ture is compatible with the design criteria, then all of this data is given back to the program manager as a Stage I report. This provides the basis for a recycling, and the Stage I activites begin anew with another staff meeting for redefinition of the program. That pretty well takes care of Stage I.

At the same time this Stage I activity is going on, the evaluator is looking around in the field to see what is actually going on there. Part of this is compatibility testing. He wants to pinpoint conflicts in facilities, use of media and so forth. Of particular importance are conflicts of space, and human resources. The other part of the fieldwork is Stage II, which is the congruence testing part of evaluation, It does no good to have the best of blueprints, if the staff are doing what they please out in the field.

Congruence testing is the comparison of some observed aspect of the program in the field, with the standard provided by the program design. Thus we have the rather elementary situation of the one independent sample research design. We derive the norm or hypothetical distribution from the standard, and the observed distribution reflects what is happening in the field. When we find that the teacher is individualizing instruction about 10% of her time, and the program design stipulates that she should be spending about 60% of her time in individual interaction, the program is off-target.

The evaluator proceeds item by item through the program design considering each variable for a congruence test. His decision on which variables to test is based on (a) considerations of researchability, and (b) the possibility of significant discrepancies being uncovered. These criteria are introduced because of the limited re-

sources available to the evaluator. A tradeoff is effected between those aspects of the program easiest to look at, and those aspects most important or most likely to be amiss.

We are thus trying to find problems in the program. As anyone who has undertaken educational research knows, it does no good to find insignificant differences across treatment levels if the lack of effects cannot be attributed to some specific failure in the program. The only decision rule for an aggregate statement of "no effects" is a cutback in program resources throughout the relevant range. On the other hand, a specific statement of "no effects due to a malfunction of component" is the basis for program change and improvement.

When the evaluator has completed his study of discrepancies between program operation and design, he again reports the findings to the program manager. This decision-maker can either make changes in the program operation, or else take the discrepancy information back to a staff definition meeting, and change the program design. By this means, we see how the rationally managed program proceeds to equalize program operation and design. In Stage I the program blueprint is ever refined, and in Stage II congruence between the standard and operation is ever increased. There is a constant interplay between the two.